

RECEIVED

AUG 18 2000

TC 1500 MAIL ROOM

ENTERED

1638

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/360,685A

DATE: 08/15/2000

TIME: 16:07:04

Input Set : A:\Chir0157.app

Output Set : N:\CRF3\08152000\I360685A.raw

```

3 <110> APPLICANT: Covacci, Antonello
4   Bugnoli, Massimo
5   Telford, John
6   Macchia, Giovanni
7   Rappuoli, Rino
9 <120> TITLE OF INVENTION: Helicobacter Pylori CAI Antigen Proteins Useful For
10  Vaccines And Diagnostics
12 <130> FILE REFERENCE: CHIR0157
14 <140> CURRENT APPLICATION NUMBER: 09/360,685A
15 <141> CURRENT FILING DATE: 1999-07-26
17 <150> PRIOR APPLICATION NUMBER: 08/471,491
18 <151> PRIOR FILING DATE: 1995-06-06
20 <160> NUMBER OF SEQ ID NOS: 10
22 <170> SOFTWARE: PatentIn Ver. 2.1
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 27
26 <212> TYPE: DNA
27 <213> ORGANISM: Artificial Sequence
29 <220> FEATURE:
30 <223> OTHER INFORMATION: Description of Artificial Sequence: Novel
31   Sequence
33 <400> SEQUENCE: 1
34 gcaagcttat cgatgtcgac tcgagct
37 <210> SEQ ID NO: 2
38 <211> LENGTH: 3960
39 <212> TYPE: DNA
40 <213> ORGANISM: Helicobacter pylori
42 <400> SEQUENCE: 2
43 aaaaagaaag gaagaaaatg gaaatacaac aaacacaccg caaaatcaat cgccctctgg 60
44 tttctctcgc tttagtagga gcattagtca gcatcacacc gcaacaaagt catgccgcct 120
45 ttttcacaac cgtgatcatt ccagccattg ttgggggtat cgctacaggc accgctgtag 180
46 gaacgggtctc agggcttctt agctgggggc tcaaaacaagc cgaagaagcc aataaaaccc 240
47 cagataaacc cgataaagtt tggcgcattc aagcaggaag aggccttaat gaattcccta 300
48 acaagggaata cgacttatac agatcccttt tatccagtaa gattgatgga ggttgggatt 360
49 gggggaatgc cgctaggcat tattgggtca aaggcgggca acagaataag ctgaagtgg 420
50 atatgaaaga cgctgtaggg acttatacct tatcagggct tagaaacttt actggtgggg 480
51 atttagatgt caatatgcaa aaagccactt tacgcttggg ccaattcaat ggcaattcct 540
52 ttacaagcta taaggatagt gctgatcgca ccacgagagt ggatttcaac gctaaaaata 600
53 tctcaattga taattttgta gaaatcaaca atcgtgtggg ttctggagcc gggaggaaa 660
54 ccagctctac ggttttgact ttgcaagctt cagaagggat cactagcgat aaaaacgctg 720
55 aaatttctct ttatgatggt gccacgctca atttggett c aagcagcggt aaattaatgg 780
56 gtaatgtgtg gatgggccgt ttgcaatacg tgggagcgta ttggcccct tcatacagca 840
57 cgataaacac ttcaaaagta acaggggaag tgaattttta ccacctcact gttggcgata 900
58 aaaacgcccgc tcaagcgggc attatcgcta ataaaaagac taatattggc acactggatt 960
59 tgtggcaaaag cgccgggtta aacattatcg ctctccaga aggtggctat aaggataaac 1020
60 ccaataatac cccttctcaa agtgggtgcta aaaacgacaa aatgaaagc gctaaaaacg 1080
61 acaacaaga gagcagtcaa aataatagta acactcaggt cattaacca cccaatagtg 1140

```

## RAW SEQUENCE LISTING

DATE: 08/15/2000

PATENT APPLICATION: US/09/360,685A

TIME: 16:07:04

Input Set : A:\Chir0157.app

Output Set: N:\CRF3\08152000\I360685A.raw

```

62 cgcaaaaaac agaagttcaa cccacgcaag tcattgatgg gccttttgcg ggcggaacaa 1200
63 acacggttgt caatatcaac cgcacgaaca ctaacgctga tggcagcatt agagtgggag 1260
64 ggttttaaac ttctcttacc accaatgcgg ctcatattgca tatcggaaca ggcggtgtca 1320
65 atctgtccaa tcaagcgagc gggcgctctc ttatagtggg aaatctaact gggaatatca 1380
66 ccgttgatgg gcctttaaga gtgaataatc aagtgggtgg ctatgctttg gcaggatcaa 1440
67 gcgcgaattt tgagttaag gctggtacgg atacaaaaaa cggcacagcc acttttaata 1500
68 acgatattag tctgggaaga tttgtgaatt taaaggtgga tgctcataca gctaatttta 1560
69 aaggtattga tacgggtaat ggtggtttca acaccttaga ttttagtggc gttacagaca 1620
70 aagtcaatat caacaagctc attacggctt ccactaatgt ggccgttaaa aacttcaaca 1680
71 ttaatgaatt gattgttaaa accaatggga taagtgtggg ggaatatact cattttagcg 1740
72 aagatatagg cagtcaatcg cgcacgaaca ccgtgcgttt ggaaactggc actaggtcac 1800
73 ttttctctgg ggtgtgtaaa tttaaaggtg gcgaaaaatt ggttatagat gagttttact 1860
74 atagcccttg gaattatttt gacgctagaa atattaaaaa tgttgaaatc accaataaac 1920
75 ttgcttttgg acctcaagga agtccttggg gcacatcaaa acttatgttc aataatctaa 1980
76 ccctaggtca aaatgcggtc atggattata gccaattttc aaatttaacc attcaagggg 2040
77 atttcatcaa caatcaaggc actatcaact atctggtccg aggtgggaaa gtggcaacct 2100
78 taagcgttag caatgcagca gctatgatgt ttaataatga tatagacagc gcgaccggat 2160
79 tttaaaaacc gctcatcaag attaacagcg ctcaagatct cattaaaaat acagaacatg 2220
80 ttttattgaa agcgaaaaatc attggttatg gtaatgtttc tacaggatcc aatggcatta 2280
81 gtaatgttaa tctagaagag caattcaaaag agcgcttagc cctttataac aacaataacc 2340
82 gcatggatac ttgtgtggtg cgaaatactg atgacattaa agcatgcggt atggctatcg 2400
83 gcgatcaaa gcatggtgaac aaccttgaca attacaagta tcttatcggt aaggcatgga 2460
84 aaaaatatag gatcagcaaa acagctaatg gctctaaaaa ttcggtgtat tatttaggca 2520
85 attctacgcc tactgagaat ggtggcaata ccacaaattt acccacaac accactagca 2580
86 atgcacgttc tgccaacaac gcccttgcac aaaacgctcc ttctgctcaa cctagtgtca 2640
87 ctctcaatth agtcgctatc aatcagcatg attttggcac tattgaaagc gtgtttgaat 2700
88 tggctaaccg ctctaaagat attgacacgc tttatgctaa ctcaggcgct caaggcaggg 2760
89 atctcttaca aaccttattg attgatagcc atgatgcggg ttatgccaga aaaaatgattg 2820
90 atgctacaag cgctaattga atcaccgaag aattgaatac ggccactacc actttaaaaca 2880
91 acatagccag tttagagcat aaaaccagcg gcttacaac tttgagcttg agtaatgcga 2940
92 tgattttaaa ttctcgttta gtcaatctct ccaggagaca caccacacat attgactcgt 3000
93 tcgccaacag cttacaagct ttaaaagacc aaaaattcgc ttcttttaga agcgcgccag 3060
94 aagtgttgta tcaatttgcc cctaaatatg aaaaacctac caatgttttg gctaacgcta 3120
95 ttgggggaac gagcttgaat aatggctcta acgcttcatt gtatggcaca agcgcgggcg 3180
96 tagacgctta ccttaacggg caagtggaa gcaattgtgg cggttttggg agctatggtt 3240
97 atagctcttt taataatcgt gcgaactccc ttaactctgg ggccaataac actaatattg 3300
98 gcgtgtatag ccgtattttt gccaacccagc atgaatttga ctttgaagct caagggggcac 3360
99 tagggagcga tcaatcaagc ttgaatttca aaagcgctct attacaagat ttgaatcaaa 3420
100 gctatcatta cttagcctat agcgctgcaa caagagcgag ctatggttat gacttcgcgt 3480
101 ttttttagga cgcttttagt ttaaaaccaa gcgtgggtgt gagctataac catttaggtt 3540
102 caaccaactt taaaagcaac agcaccatc aagtggcttt gaaaaatggc tctagcagtc 3600
103 agcatttatt caacgctagc gctaattgtg aagcgcgcta ttattatggg gacacttcat 3660
104 acttctacat gaatgctgga gttttacaag agttcgctca tgttggctct aataacgccg 3720
105 cgtctttaa cacttttaa gtgaatgcc ctcgcaaccc tttaaatacc catgcccagag 3780
106 tgatgatggg tggggaatta aaattagcta aagaagtgtt tttgaatttg ggcgtttgtt 3840
107 atttgcaaaa ttgattttcc aatatagccc atttcgcttc caatttagga atgaggtata 3900
108 gtttctaata accgctctta aacccatgct caaagcatgg gtttgaatc ttacaaaaa 3960
111 <210> SEQ ID NO: 3
112 <211> LENGTH: 1296

```

RAW SEQUENCE LISTING                      DATE: 08/15/2000  
 PATENT APPLICATION: US/09/360,685A        TIME: 16:07:04

Input Set : A:\Chir0157.app  
 Output Set: N:\CRF3\08152000\I360685A.raw

```

113 <212> TYPE: PRT
114 <213> ORGANISM: Helicobacter pylori
116 <400> SEQUENCE: 3
117 Met Glu Ile Gln Gln Thr His Arg Lys Ile Asn Arg Pro Leu Val Ser
118   1           5           10           15
120 Leu Ala Leu Val Gly Ala Leu Val Ser Ile Thr Pro Gln Gln Ser His
121           20           25           30
123 Ala Ala Phe Phe Thr Thr Val Ile Ile Pro Ala Ile Val Gly Gly Ile
124           35           40           45
126 Ala Thr Gly Thr Ala Val Gly Thr Val Ser Gly Leu Leu Ser Trp Gly
127           50           55           60
129 Leu Lys Gln Ala Glu Glu Ala Asn Lys Thr Pro Asp Lys Pro Asp Lys
130   65           70           75           80
132 Val Trp Arg Ile Gln Ala Gly Lys Gly Phe Asn Glu Phe Pro Asn Lys
133           85           90           95
135 Glu Tyr Asp Leu Tyr Arg Ser Leu Leu Ser Ser Lys Ile Asp Gly Gly
136           100          105          110
138 Trp Asp Trp Gly Asn Ala Ala Arg His Tyr Trp Val Lys Gly Gly Gln
139           115          120          125
141 Gln Asn Lys Leu Glu Val Asp Met Lys Asp Ala Val Gly Thr Tyr Thr
142           130          135          140
144 Leu Ser Gly Leu Arg Asn Phe Thr Gly Gly Asp Leu Asp Val Asn Met
145 145           150          155          160
147 Gln Lys Ala Thr Leu Arg Leu Gly Gln Phe Asn Gly Asn Ser Phe Thr
148           165          170          175
150 Ser Tyr Lys Asp Ser Ala Asp Arg Thr Thr Arg Val Asp Phe Asn Ala
151           180          185          190
153 Lys Asn Ile Ser Ile Asp Asn Phe Val Glu Ile Asn Asn Arg Val Gly
154           195          200          205
156 Ser Gly Ala Gly Arg Lys Ala Ser Ser Thr Val Leu Thr Leu Gln Ala
157           210          215          220
159 Ser Glu Gly Ile Thr Ser Asp Lys Asn Ala Glu Ile Ser Leu Tyr Asp
160 225           230          235          240
162 Gly Ala Thr Leu Asn Leu Ala Ser Ser Ser Val Lys Leu Met Gly Asn
163           245          250          255
165 Val Trp Met Gly Arg Leu Gln Tyr Val Gly Ala Tyr Leu Ala Pro Ser
166           260          265          270
168 Tyr Ser Thr Ile Asn Thr Ser Lys Val Thr Gly Glu Val Asn Phe Asn
169           275          280          285
171 His Leu Thr Val Gly Asp Lys Asn Ala Ala Gln Ala Gly Ile Ile Ala
172           290          295          300
174 Asn Lys Lys Thr Asn Ile Gly Thr Leu Asp Leu Trp Gln Ser Ala Gly
175 305           310          315          320
177 Leu Asn Ile Ile Ala Pro Pro Glu Gly Gly Tyr Lys Asp Lys Pro Asn
178           325          330          335
180 Asn Thr Pro Ser Gln Ser Gly Ala Lys Asn Asp Lys Asn Glu Ser Ala
181           340          345          350
183 Lys Asn Asp Lys Gln Glu Ser Ser Gln Asn Asn Ser Asn Thr Gln Val
184           355          360          365

```

## RAW SEQUENCE LISTING

DATE: 08/15/2000

PATENT APPLICATION: US/09/360,685A

TIME: 16:07:04

Input Set : A:\Chir0157.app

Output Set: N:\CRF3\08152000\I360685A.raw

```

186 Ile Asn Pro Pro Asn Ser Ala Gln Lys Thr Glu Val Gln Pro Thr Gln
187      370      375      380
189 Val Ile Asp Gly Pro Phe Ala Gly Gly Lys Asp Thr Val Val Asn Ile
190 385      390      395      400
192 Asn Arg Ile Asn Thr Asn Ala Asp Gly Thr Ile Arg Val Gly Gly Phe
193      405      410      415
195 Lys Ala Ser Leu Thr Thr Asn Ala Ala His Leu His Ile Gly Lys Gly
196      420      425      430
198 Gly Val Asn Leu Ser Asn Gln Ala Ser Gly Arg Ser Leu Ile Val Glu
199      435      440      445
201 Asn Leu Thr Gly Asn Ile Thr Val Asp Gly Pro Leu Arg Val Asn Asn
202      450      455      460
204 Gln Val Gly Gly Tyr Ala Leu Ala Gly Ser Ser Ala Asn Phe Glu Phe
205 465      470      475      480
207 Lys Ala Gly Thr Asp Thr Lys Asn Gly Thr Ala Thr Phe Asn Asn Asp
208      485      490      495
210 Ile Ser Leu Gly Arg Phe Val Asn Leu Lys Val Asp Ala His Thr Ala
211      500      505      510
213 Asn Phe Lys Gly Ile Asp Thr Gly Asn Gly Gly Phe Asn Thr Leu Asp
214      515      520      525
216 Phe Ser Gly Val Thr Asp Lys Val Asn Ile Asn Lys Leu Ile Thr Ala
217      530      535      540
219 Ser Thr Asn Val Ala Val Lys Asn Phe Asn Ile Asn Glu Leu Ile Val
220 545      550      555      560
222 Lys Thr Asn Gly Ile Ser Val Gly Glu Tyr Thr His Phe Ser Glu Asp
223      565      570      575
225 Ile Gly Ser Gln Ser Arg Ile Asn Thr Val Arg Leu Glu Thr Gly Thr
226      580      585      590
228 Arg Ser Leu Phe Ser Gly Gly Val Lys Phe Lys Gly Gly Glu Lys Leu
229      595      600      605
231 Val Ile Asp Glu Phe Tyr Tyr Ser Pro Trp Asn Tyr Phe Asp Ala Arg
232      610      615      620
234 Asn Ile Lys Asn Val Glu Ile Thr Asn Lys Leu Ala Phe Gly Pro Gln
235 625      630      635      640
237 Gly Ser Pro Trp Gly Thr Ser Lys Leu Met Phe Asn Asn Leu Thr Leu
238      645      650      655
240 Gly Gln Asn Ala Val Met Asp Tyr Ser Gln Phe Ser Asn Leu Thr Ile
241      660      665      670
243 Gln Gly Asp Phe Ile Asn Asn Gln Gly Thr Ile Asn Tyr Leu Val Arg
244      675      680      685
246 Gly Gly Lys Val Ala Thr Leu Ser Val Gly Asn Ala Ala Met Met
247      690      695      700
249 Phe Asn Asn Asp Ile Asp Ser Ala Thr Gly Phe Tyr Lys Pro Leu Ile
250 705      710      715      720
252 Lys Ile Asn Ser Ala Gln Asp Leu Ile Lys Asn Thr Glu His Val Leu
253      725      730      735
255 Leu Lys Ala Lys Ile Ile Gly Tyr Gly Asn Val Ser Thr Gly Thr Asn
256      740      745      750
258 Gly Ile Ser Asn Val Asn Leu Glu Glu Gln Phe Lys Glu Arg Leu Ala

```

## RAW SEQUENCE LISTING

DATE: 08/15/2000

PATENT APPLICATION: US/09/360,685A

TIME: 16:07:04

Input Set : A:\Chir0157.app

Output Set: N:\CRF3\08152000\I360685A.raw

```

259          755          760          765
261 Leu Tyr Asn Asn Asn Asn Arg Met Asp Thr Cys Val Val Arg Asn Thr
262          770          775          780
264 Asp Asp Ile Lys Ala Cys Gly Met Ala Ile Gly Asp Gln Ser Met Val
265 785          790          795          800
267 Asn Asn Pro Asp Asn Tyr Lys Tyr Leu Ile Gly Lys Ala Trp Lys Asn
268          805          810          815
270 Ile Gly Ile Ser Lys Thr Ala Asn Gly Ser Lys Ile Ser Val Tyr Tyr
271          820          825          830
273 Leu Gly Asn Ser Thr Pro Thr Glu Asn Gly Gly Asn Thr Thr Asn Leu
274          835          840          845
276 Pro Thr Asn Thr Thr Ser Asn Ala Arg Ser Ala Asn Asn Ala Leu Ala
277          850          855          860
279 Gln Asn Ala Pro Phe Ala Gln Pro Ser Ala Thr Pro Asn Leu Val Ala
280 865          870          875          880
282 Ile Asn Gln His Asp Phe Gly Thr Ile Glu Ser Val Phe Glu Leu Ala
283          885          890          895
285 Asn Arg Ser Lys Asp Ile Asp Thr Leu Tyr Ala Asn Ser Gly Ala Gln
286          900          905          910
288 Gly Arg Asp Leu Leu Gln Thr Leu Leu Ile Asp Ser His Asp Ala Gly
289          915          920          925
291 Tyr Ala Arg Lys Met Ile Asp Ala Thr Ser Ala Asn Glu Ile Thr Lys
292          930          935          940
294 Gln Leu Asn Thr Ala Thr Thr Thr Leu Asn Asn Ile Ala Ser Leu Glu
295 945          950          955          960
297 His Lys Thr Ser Gly Leu Gln Thr Leu Ser Leu Ser Asn Ala Met Ile
298          965          970          975
300 Leu Asn Ser Arg Leu Val Asn Leu Ser Arg Arg His Thr Asn His Ile
301          980          985          990
303 Asp Ser Phe Ala Lys Arg Leu Gln Ala Leu Lys Asp Gln Lys Phe Ala
304          995          1000          1005
306 Ser Leu Glu Ser Ala Ala Glu Val Leu Tyr Gln Phe Ala Pro Lys Tyr
307          1010          1015          1020
309 Glu Lys Pro Thr Asn Val Trp Ala Asn Ala Ile Gly Gly Thr Ser Leu
310 1025          1030          1035          1040
312 Asn Asn Gly Ser Asn Ala Ser Leu Tyr Gly Thr Ser Ala Gly Val Asp
313          1045          1050          1055
315 Ala Tyr Leu Asn Gly Gln Val Glu Ala Ile Val Gly Gly Phe Gly Ser
316          1060          1065          1070
318 Tyr Gly Tyr Ser Ser Phe Asn Asn Arg Ala Asn Ser Leu Asn Ser Gly
319          1075          1080          1085
321 Ala Asn Asn Thr Asn Phe Gly Val Tyr Ser Arg Ile Phe Ala Asn Gln
322          1090          1095          1100
324 His Glu Phe Asp Phe Glu Ala Gln Gly Ala Leu Gly Ser Asp Gln Ser
325 1105          1110          1115          1120
327 Ser Leu Asn Phe Lys Ser Ala Leu Leu Gln Asp Leu Asn Gln Ser Tyr
328          1125          1130          1135
330 His Tyr Leu Ala Tyr Ser Ala Ala Thr Arg Ala Ser Tyr Gly Tyr Asp
331          1140          1145          1150

```

VERIFICATION SUMMARY                      DATE: 08/15/2000  
PATENT APPLICATION:    US/09/360,685A        TIME: 16:07:05  
  
Input Set : A:\Chir0157.app  
Output Set: N:\CRF3\08152000\I360685A.raw